

# Data Warehouse Crash Course

Eshant Garg

Data Engineer, Architect, Advisor

[eshant.garg@gmail.com](mailto:eshant.garg@gmail.com)



# Introduction



Data Warehouse basics

Benefits over Transactional Database

Dimensional Modeling

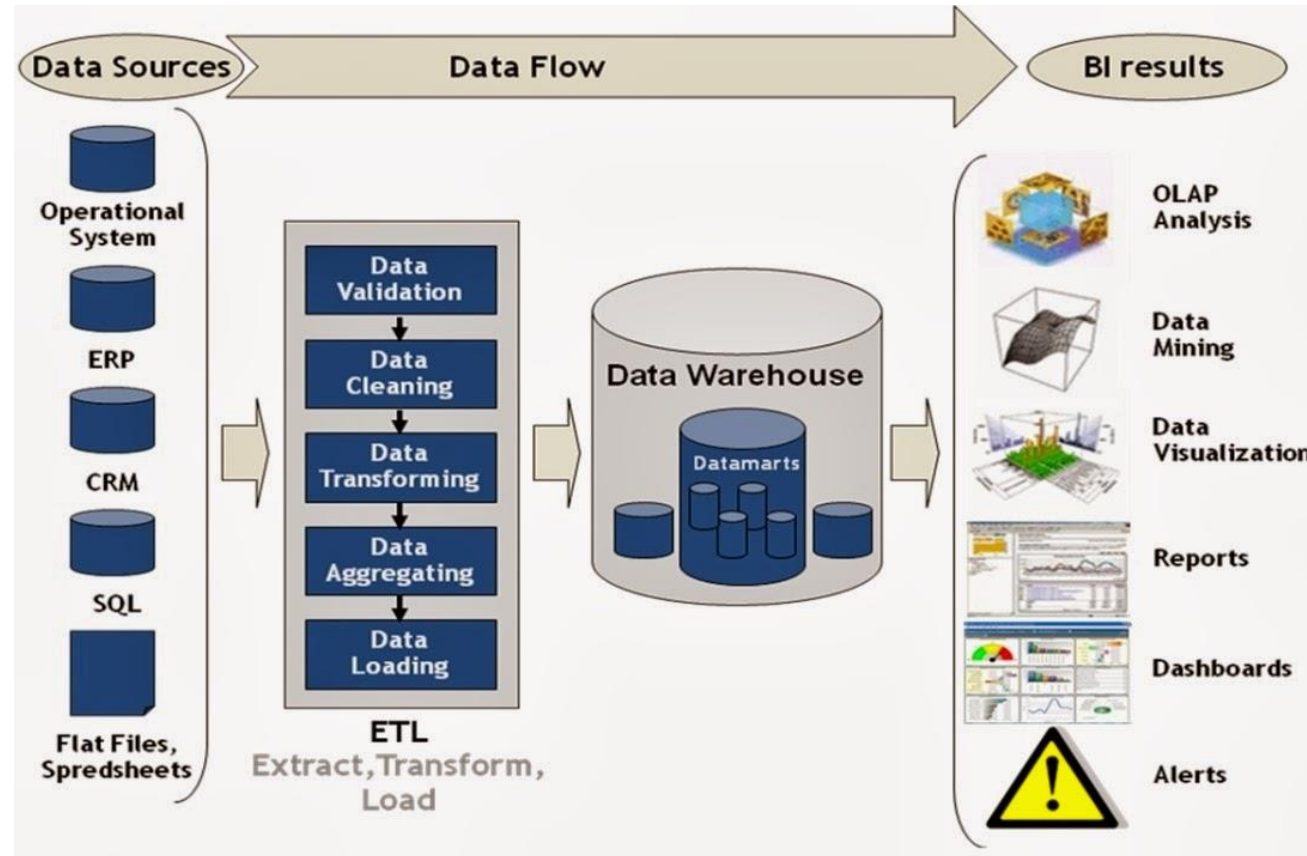
Facts and Fact Table

Dimensions and Dimension Table

Star vs Snowflake Schema

Why we need Data warehouse?

# What is Data Warehouse



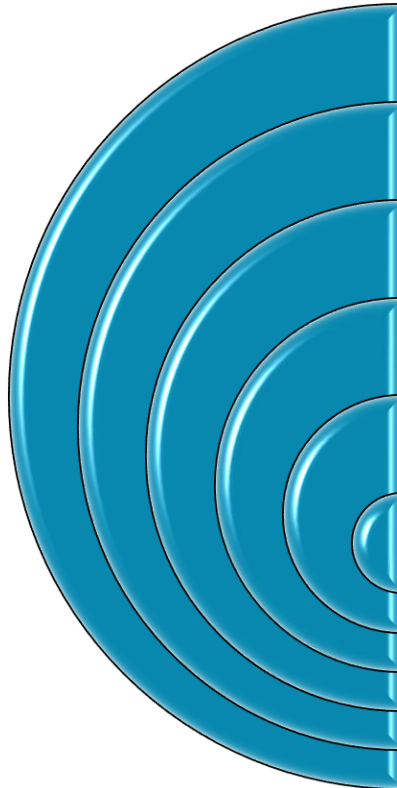
# Concerns About a Data Warehouse

Extra piece of software that needs maintenance

The data is already present in the operational systems

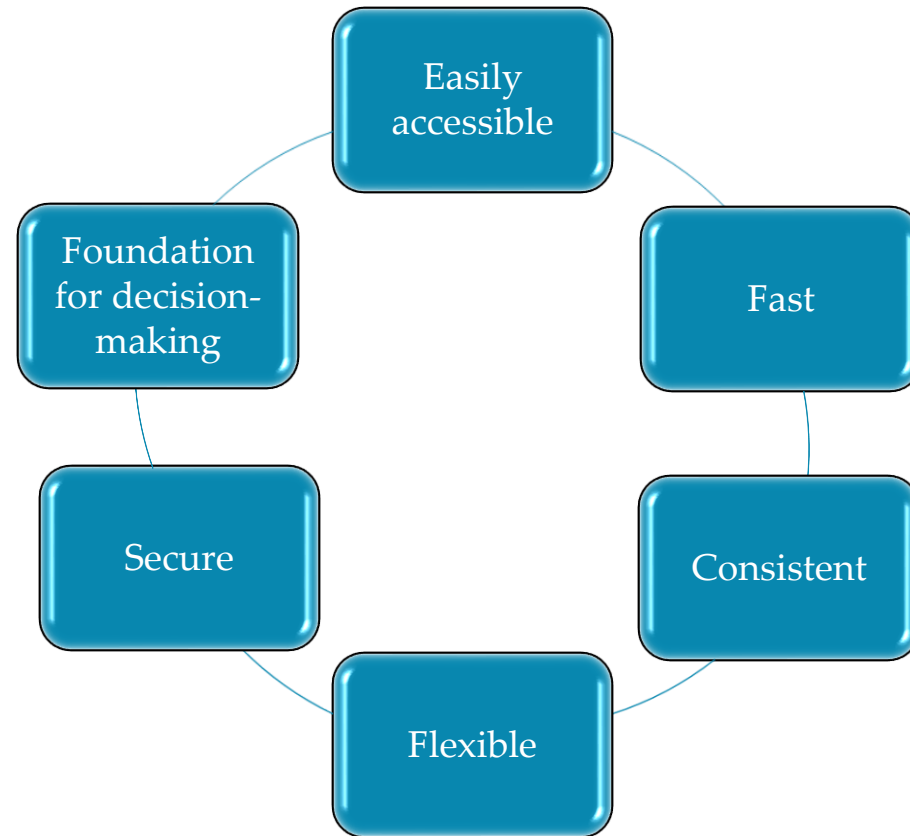
However, extracting information from operational data is complicated

# Problems a Data Warehouse Can Solve



We have so much data, but we can't make anything of it
I only want to know what is important
We need to slice and dice the data
Business people need to access data easily
Numbers between departments don't match and we don't know who's right
We want people to make decisions based on facts

# Ideal Data Warehouse Solution



# Responsibilities of a Data Warehouse Designer

Understand the  
business users goal  
and objectives

Deliver accurate,  
trustworthy and  
relevant  
information

Sustain the DW  
environment



# OLTP vs OLAP

## OLTP

- Many small transaction
- Current data
- Used to run the business
- Highly detailed
- Typically in the GB scale
- Processing performance limit

## OLAP

- Low volume but complex queries
- Historic, non-volatile data
- Used to analyze the business
- Consolidated and summarized
- TB and above scale
- No limit, pause/resume compute

# Dimensional Modeling

**Database design method optimized for data warehouse solutions**


**Popular technique because it addresses two important requirements**

1. Deliver data in an understandable format
2. Deliver fast query performance

**Key word is “simplicity”**

**Sustain the DW environment**

# Elements of Dimensional Model



Facts	<ul style="list-style-type: none"><li>• The measurements or metrics or facts from your business process</li></ul>
Dimensions	<ul style="list-style-type: none"><li>• For providing the context of a business process event</li></ul>
Attributes	<ul style="list-style-type: none"><li>• The various characteristics of a dimension</li></ul>
Star Schema	<ul style="list-style-type: none"><li>• And/or OLAP cubes</li></ul>

"We sell Cake and other Products in various locations and measure our achievements over time."

# Facts and Fact Tables



**Fact** = a business measure

- Sales
- Profit
- Volume
- Number of transactions

-	-	-
-	-	-
-	-	-
-	-	-

**Fact table** = table that stores the performance measurements resulting from an organization's business process events

# Facts and Fact Tables



Facts answer questions like:

What are we doing? (sell, buy, count)

What do we want to achieve? (more sales, bigger profit)

1 row in the fact table is 1 measurement in real life

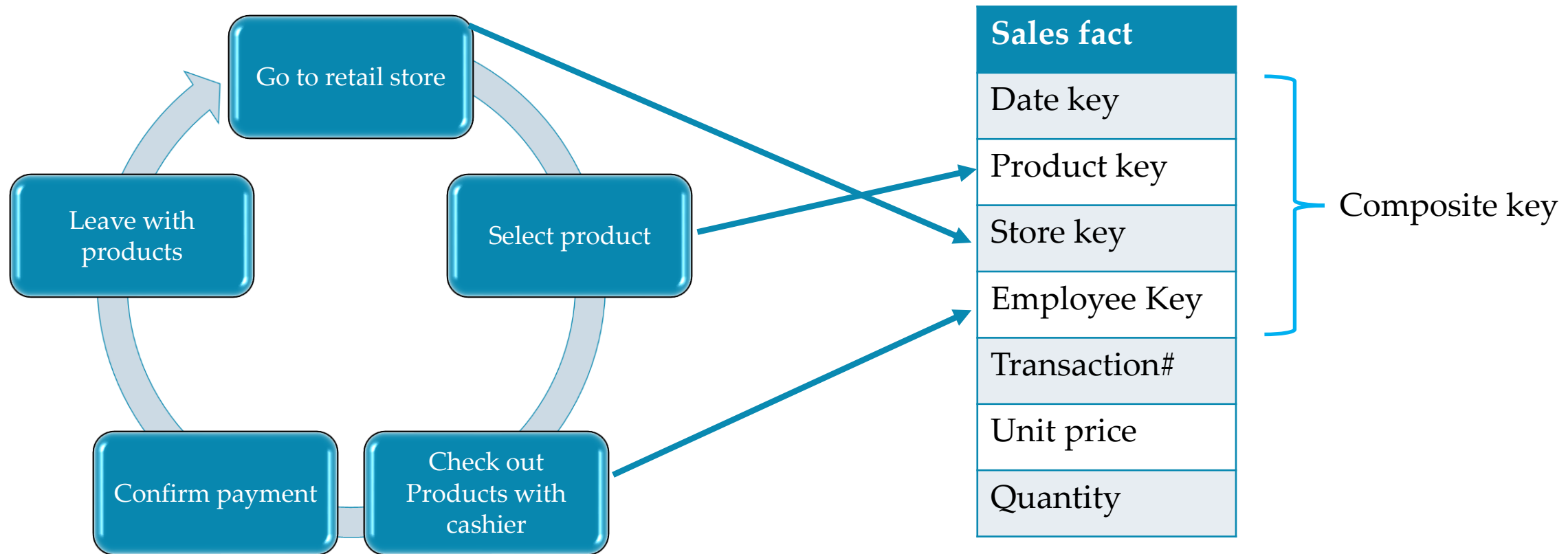
Fact columns in a fact table should be additive

Facts make sense in combination with dimensions

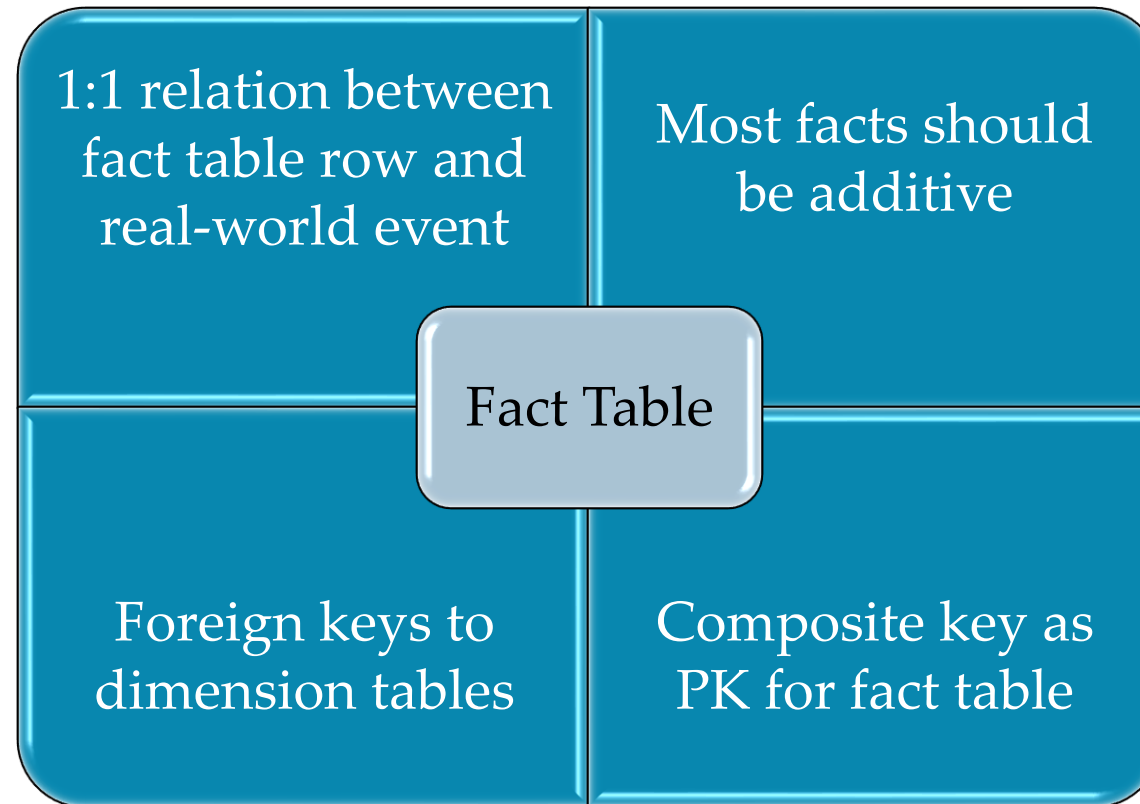
Linked with foreign keys

Dat/Time dimension is present in most data warehouses

# Example of a Fact Table

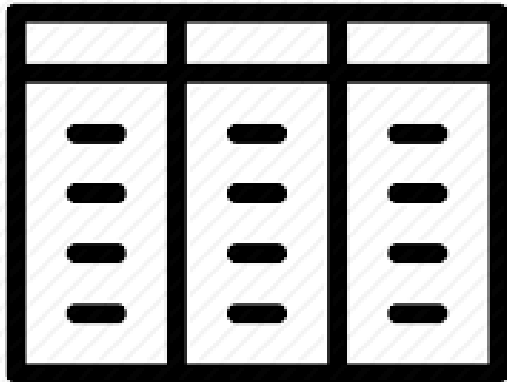


# Characteristics of Fact Tables





# What are Dimensions?



-	-	-
-	-	-
-	-	-
-	-	-

Companions to a fact table

Textual context associated with a business  
process measurement event

# Questions Answered by Dimension Tables



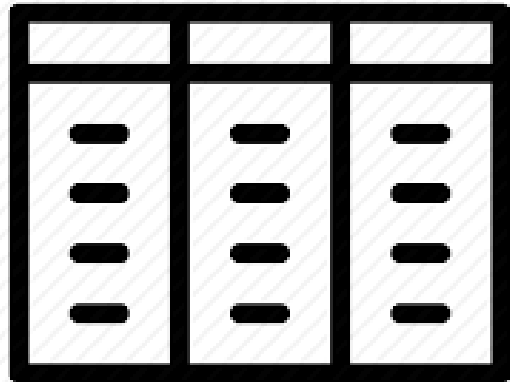
# Example of a Dimension Table



Sales fact
Date key
Product key
Store key
Employee Key
Transaction#
Unit price
Quantity

Product Dimension
Product key
Product name
Brand name
Category name
Subcategory name
Package type
Package size
Weight
Weight unit of measure

# Characteristics of Dimension Tables

A diagram of a dimension table structure. It consists of a table with 3 columns and 5 rows. The top row is shaded with diagonal lines and contains three horizontal bars, representing a primary key. The remaining four rows are also shaded with diagonal lines and each contains three horizontal bars, representing data rows. The table is enclosed in a thick black border.

—	—	—
—	—	—
—	—	—
—	—	—

## **No limit for the number of attribute in a dimension table**

Common to have tables with 50 to 100 attributes

Some dimension tables have only a handful of attributes

## **Have fewer rows than fact tables**

But can be much wider

## **Defined by a single primary key**

Basis for referential integrity with the fact table

## **Denormalized**

Flattened many-to-one relationships within a single dimension table

# Example of a Dimension Table

## Characteristics

Simple

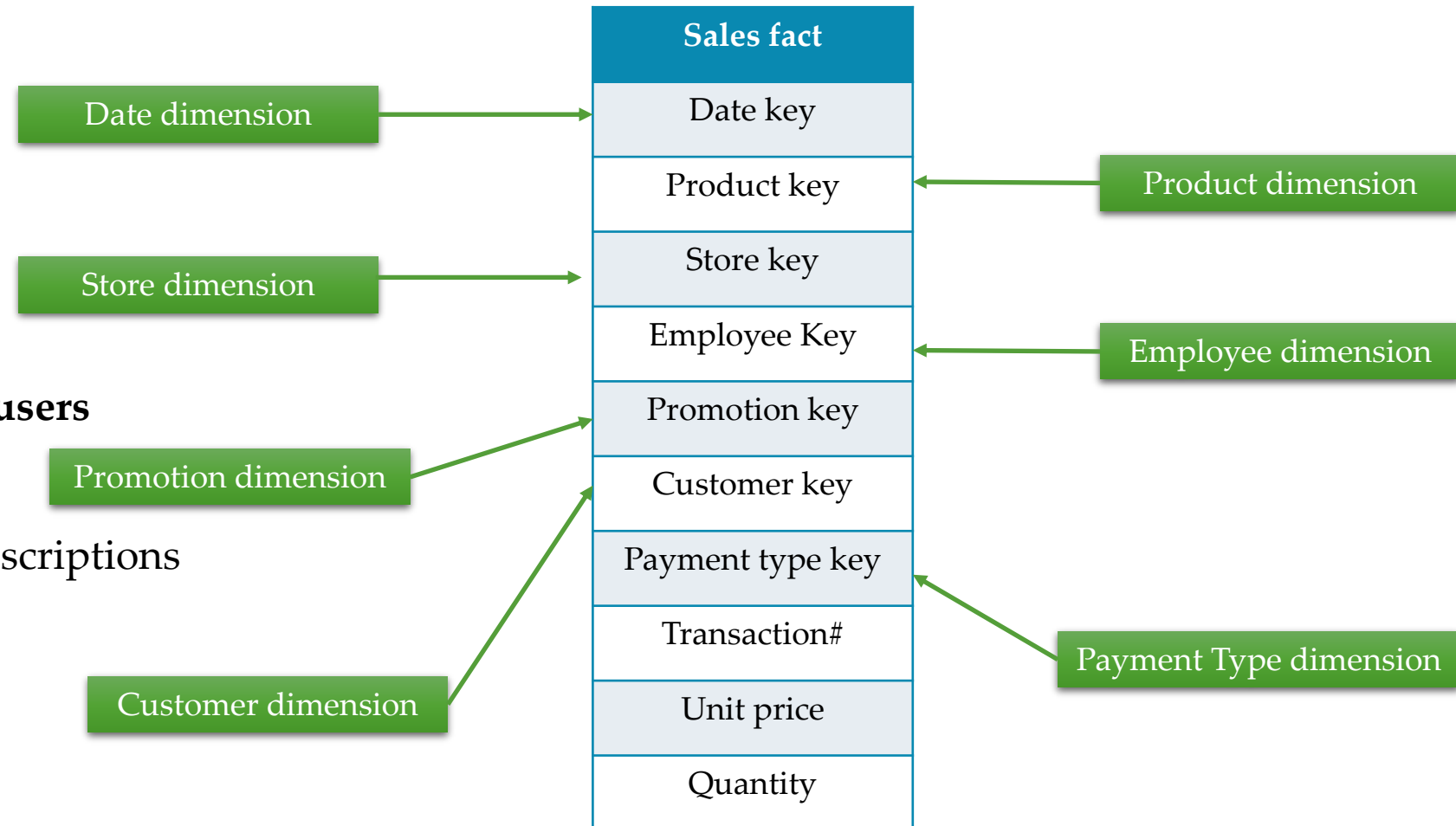
Symmetric

## Highly-recognizable by business users

Reduced number of tables

Use of meaningful business descriptions

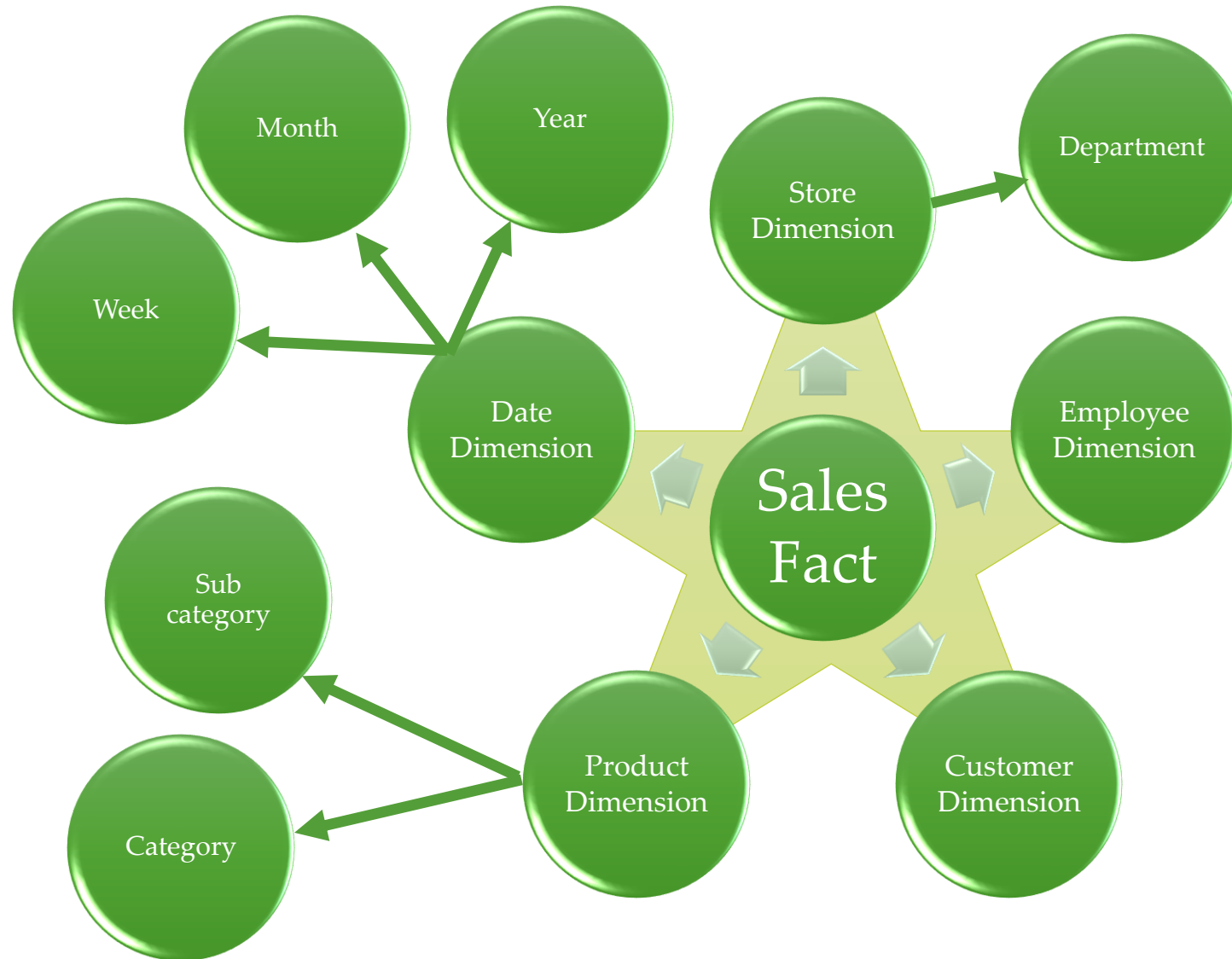
## Extensible



# Dimensional Model as a Star Schema



# Dimensional Model as a Snow flake



# Summary

What is Data Warehouse

Why we need Data Warehouse

Data Warehouse (OLAP) vs Transactional Database (OLTP)

Dimensional Modeling

Facts and Fact Tables

Dimensions and Dimension tables

Star vs Snowflake Schema

